Challenge to Judgement

Conference Paper · October 2019

Charles MacRobert
Stellenbosch University

INTERNATIONAL SOCIETY FOR SOIL MECHANICS AND GEOTECHNICAL ENGINEERING

The paper was published in the proceedings of the 17th African Regional Conference on Soil Mechanics and Geotechnical Engineering and was edited by Prof. Denis Kalumba. The conference was held in Cape Town, South Africa, on October 07-09 2019.

Proceedings of the 17th African Regional Conference on Soil Mechanics and Geotechnical Engineering. 7, 8 & 9 October 2019 – Cape Town

Challenge to Judgement

C.J. MacRobert
Stellenbosch University, Stellenbosch, South Africa

ABSTRACT: Despite significant advances in technical knowledge, failures of geotechnical structures still take place. Often this is due to large epistemic uncertainties in our understanding of ground conditions and a failure to apply correct judgement. To better understand judgement, participants active in the geotechnical field were asked to predict the safety of five typical geotechnical problems. Problems were presented to elicit potential mistakes in judgement. Findings showed in general that it is difficult to make correct intuitive decisions for complex problems under time constraints, better decisions could be made for simpler problems and stating reasons for particular decisions resulted in improved judgments. Whilst better decisions were often made by more experienced practitioners this was not consistently the case. Confidence in one's answer was found to be a poor predictor of good judgement. This illustrated that judgement is fallible despite experience or confidence.

1 INTRODUCTION

The title of this paper comes from the economist Paul Samuelson (1974) seminal paper "Challenge to Judgement". In his paper, Samuelson (1974) questioned whether any active fund manager could ever outperform the average market. This led to the index fund or basket of stocks chosen to passively track the average market. As engineers we should, from time to time, take stock of our judgement. Does it work and how can we improve? One would hope we can make better predictions on the performance of structures than economists can the performance of stocks.

Vick (2002) defines 'engineering judgement' as the ability to integrate evidence from different sources, including personal experience, in coming to a decision. An analysis of foundation failures by Sowers (1993) suggested that 88% of failures were due to 'human shortcomings' and the rest due to a lack of technological understanding. The prevalence of failures due to 'human shortcomings' is unsettling in the field of geotechnical engineering where 'engineering judgement' is often relied upon.

Behavioural psychology suggests that humans are prone to making wrong intuitive judgements (Kahneman 2012). However, intuitive expertise has been shown to develop in some environments where there is a stable relationships between identifiable cues and events, and individuals have sufficient time to learn them (Kahneman and Klein 2009). Experts develop rules of thumb, which although potentially fallible guide decision making (Koen 2003).

Fundamental to engineering design is safety. A design is proposed and then the likelihood of failure is determined by subjecting the design to analytical and/or physical testing (Petroski 1985). Engineers, within a given field, should develop sufficient judgment to have an intuitive feel for a design that is likely to be safe. This paper is limited to understanding whether geotechnical practitioners have an intuitive feel for safety, as this is arguably one of the most important aspect of judgement.

To investigate the use of judgement, five typical geotechnical problems were posed to practitioners of varying experience (in terms of years and projects):

1. A tailings dam that participants were unaware had failed.
2. A tailings dam that participants were unaware had performed satisfactorily.
3. A hypothetical waste capping system that was potentially unstable.
4. A foundation for a tall chimney that had performed satisfactorily.
5. A gabion retaining wall that participants were unaware had failed.

Participants were required to make predictions about the safety of each structure based on different amounts of information and after completing different tasks that could aid decision making. Lommler (2012) suggests that most engineers find it difficult to integrate more than 2 to 3 controlling factors in solving geotechnical problems. Complex problems should be broken down into smaller parts that can be solved easily as de Mello (1977) famously illustrated with the legend of the triplet Horatii and Curiatii brothers in his Rankine lecture.

The geotechnical problems were posed in different studies over a three-year period to different practitioners known to be active in the respective areas based on attendance of industry related workshops, conferences and lectures. All participants were made aware of the type of problem that would be presented prior to participation. Self-selection could therefore take place with only respondents comfortable with the stated issues participating.

Study 1 was based on Problem 1. Study 2 based on Problems 2 and 3, and Study 3 was based on Problems 4 and 5. Prior to tackling any of the problems, participants provided answers to baseline questions to establish their experience.

2 METHOD

2.1 *Study 1*

The problem in the first study was based on a case study presented by Blight (2010) for the failure of a tailings dam operated by an unnamed small scale alluvial diamond miner. The problem was presented incrementally as a one on one paper-based exercise. First participants were given a brief description of the facility, grading of materials, results from a single characteristic cone penetration test and a representative outer geometry. Participants were asked to list potential stability concerns, steps taken to address these, suggest a factor of safety and verbal term to express the probability of failure.

After this initial task, participants were asked to sketch a limit equilibrium cross section. A blank outer geometry was provided and participants had to sketch the internal geometry of materials, suggest material parameters, sketch the location of the phreatic surface and sketch the location of the critical slip circle. Participants were provided with some common correlations to interpret the cone penetration test. Participants had to again suggest a factor of safety and verbal term to express the probability of failure.

The next task was to input strength parameters into a template limit equilibrium model based on the analysis presented by Blight (2010). Participants were not told what the basis for the template was. This template consisted of three internal regions (which participants had to suggest strength parameters for) and a representative phreatic surface. Slip circle entry points were set along the top of the facility and exit points along the front face. Participants then discussed what changes they would make to the model and how this would influence the predicted factor of safety. After this discussion participants were asked to state a final factor of safety that integrated all the evidence and suggest a verbal term for the likelihood of failure (Table 1).

Table 1. Verbal terms suggested in Study 1

Verbal term
probable
low chance
very high chance
likely
even chance
almost improbable
very improbable
unlikely
very low chance
medium chance
almost certain

This problem therefore required participants to work through a number of tasks that would aid their final decision. Whilst, such an approach would be representative of actual practice, participants completed the tasks with a 30-minute countdown timer running. Participants could only refer to material provided and were not permitted to use calculators. Participants could ask clarifying question with answers limited to not give away the solution. Further details of the study can be found in MacRobert (2018b).

2.2 *Study 2*

The second study was an A/B study, in which two questions were presented to two groups but with different amounts of information provided to each. The first question pair pertained to the safety of a proposed rate of rise for a tailings dam over a 20-year period. The second question pair required the assessment of a proposed waste cover system.

The rate of rise question pair was based on a field study presented in MacRobert and Blight (2013) for a platinum tailings dam that was raised safely at a rate or rise of 2.3 m/year. Both groups were provided with material grading (10 % clay-sized particles, 70 % silt-sized particles and 20 % sand-sized particles) but the second group (Group B) was also told that the coefficient of consolidation was 150 m^2/year.

Both the grading and coefficient of consolidation would suggest that a rate of rise of around 2 m/year would be satisfactory (Robinson 2008). However, a safe rate of rise is also based on factors such as unusual weather, poor operational management, foundation instability, overtopping, seepage and structural defects (Azam & Li 2010). Safety in this question pair was therefore dependent on a large number of factors, and although key factors were provided, the amount of information provided did not substantially change between the two paired questions.

The second question pair was based on a hypothetical waste cover design for a 20 m high facility. The suggested design was a two-component cover, consisting of a suitably anchored geomembrane unto which a soil drainage layer would be placed. The soil drainage layer would be placed bottom up on a 1 to 3 slope (18.4 °). A geomembrane with a residual soil-geomembrane interface friction angle of 22 ° was proposed. Both groups were given the slope angle but the first group (Group A) was also given the interface friction angle.

Typically in South African practice, covers steeper than 1 in 4 are avoided as they have a greater propensity for stability issues (Nortje 2018). This should raise suspicions of the design in both of the paired questions as the stated slope angle is 1 to 3. The greatest source of uncertainty in the stability of covers, is the soil-geomembrane interface friction angle. For Group A, a typical interface friction angle was given and it was greater than the slope angle. Based on the model proposed by Koerner and Hwu (1991) this would result in a factor of safety of 1.29 and probability of failure of 0.14 (based on a 20 000 Monte Carlo simulation with typical variation of parameters). Fewer factors control safety in this question pair and the additional information provided to Group A has a significant influence on stability.

This second study was presented as an online questionnaire. No tasks apart from reading the provided information and predicting a likelihood of failure was required. Participants were asked not to carry out any calculations. This was therefore removed from actual practice as decisions were based purely on intuition.

Answers to these questions were selected from the calibration table given in Table 2. A calibration table was given as people (including engineers) often find it difficult to convert verbal terms into numerical probabilities (MacRobert 2018a). This table was developed by reviewing research presented by Reagan et al. (1989) and sample calibration tables in Vick (2002) and USBR-USACE (2015).

Table 2. Possible selections for Study 3

Verbal term	Numerical probability of failure
Almost improbable	0.001
Very improbable	0.01
Very unlikely	0.1
Unlikely	0.2
Even chance	0.5
Probable	0.8
Likely	0.9
Very high chance	0.99
Almost certain	0.999

2.3 *Study 3*

The third study was an A/B study, in which the same two questions were presented to two groups but with each group required to carry out different tasks.

One question pertained to the foundation for a tall chimney. The question was based on a case study presented by Davie & Lewis (1988) of a 107 m high concrete chimney (total load = 31 000 kN, wind moment at base = 65 000 kNm and eccentricity = 2.1 m) founded on a 16.5 m square pad (applied pressure = 115 kPa), underlain by medium to dense fine to coarse sand (Average SPT N = 40). The chimney foundation performed adequately and settled roughly 6 mm. Participants were presented with a brief description of the chimney and proposed foundation, SPT N log and site stratigraphy. Based on this, participants had to select a probability of complete foundation failure from Table 2.

The second question was based on a case study for a gabion retaining wall that failed (Nowatzki and Wrench 1988). The day following construction, the wall failed by rotation through a soft clay layer. No geotechnical investigation was carried out prior to construction. However, following the failure boreholes were drilled and laboratory work undertaken. Participants were presented with a description, typical cross-section, SPT N log and site stratigraphy and asked to predict the likelihood of complete failure from Table 2.

Whilst the same information was provided for both questions to both groups, the order in which the questions were presented was different. For Group A, the foundation question was presented first and participants asked to make a prediction based purely on a read through the information. The retaining wall question was then given, however, this time participants were requested to detail their reasoning (descriptive, numerical or sketches) before making a prediction. Participants were not allowed to look at reference material or use calculators. For Group B, the retaining wall question came first (with the decision based only on a read through) and the foundation question second (with the decision based on reasoning).

This third study was presented as a one on one paper-based exercise with participants able to ask clarifying questions with answers limited to not give away solutions. No time limit was stipulated and candidates were encouraged to take as much time as required. It would be expected that such an approach would lead to the best judgement decisions.

3 RESULTS

3.1 *Study 1*

The interpretation of results from Study 1 was difficult as respondents gave both a factor of safety and a verbal term for the likelihood of failure. Interpreting the results based on factors of safety seemed attractive as this was a numeric quantity that could be compared to the common limits for acceptable stability (FS = 1.3 for monitored slopes and FS = 1.5 for unmonitored slopes). However, these limits are debatable and the slope in question failed. Consequently, responses were grouped into five categories based on stated verbal terms. Respondents indicating that the likelihood of failure was ‘almost certain’ or had a ‘very high chance’ were judged to have predicted the risk of failure to be extremely high. Respondents indicating failure was ‘likely’ or ‘probable’ were judged to see the probability of failure to be high.

‘Even chance’ and ‘medium chance’ were grouped together as even. ‘Unlikely’ and ‘low chance’ responses were grouped as indicating respondents saw a low likelihood of failure. ‘Very improbable’, ‘almost improbable’, ‘very low chance’ and ‘very unlikely’ were all taken as indicating an extremely low probability of failure.

As the facility failed one would hope that a large percentage of the 57 individuals who participated would see the chance of failure as being extremely high, especially as the majority of participants are actively involved in the tailings space (7 were final year civil engineering students who performed well in course work). However, this was not the case as shown in Figure 1. The bottom panel shows that only 14% were convinced failure was imminent and a further 11% saw the probability of failure as high. However, 75% of participants did not see failure as imminent.

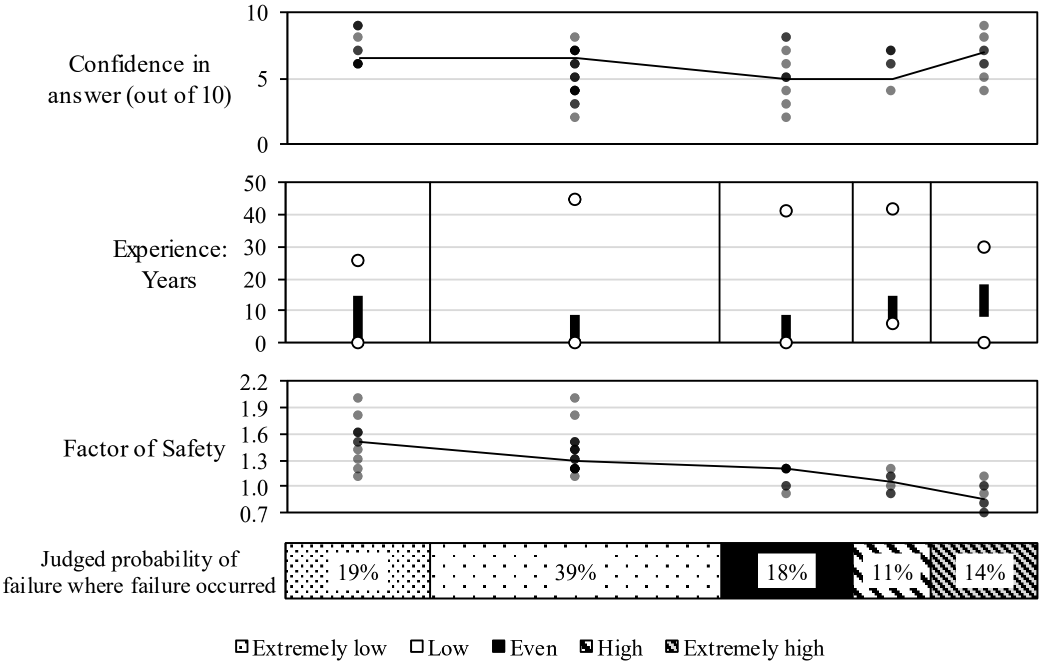

Figure 1. Results from final assessment of stability in Study 1

The median FS for extremely low and low correspond to limit FS for unmonitored (FS = 1.5) and monitored slopes (FS = 1.3) respectively. Whilst there was a large degree of scatter (darker points correspond to participants selecting identical FS) no FS below 1.0 were selected for these two categories. For even chance, the median FS was 1.2, only one FS was selected below 1.0. For the high category the median FS was 1.05, with others ranging between 1.2 and 0.9. For the extremely high category the median FS was 0.85, with others ranging between 0.7 and 1.1. This correspondence between FS and verbal terms suggests that the categorisation of responses reflected participants views of the likelihood of failure.

Whilst these results could be seen to paint a poor picture of engineering judgement the exercise was designed to elicit potential mistakes.

Behavioural psychology has shown that anchor values can have a controlling influence over predictions such that other cues are ignored (Sherif et al. 1958). Forty seven percent (47%) of respondents suggested $1.2 \leq FS \leq 1.3$ and twenty three percent (23%) selected $1.4 \leq FS \leq 1.6$. The prevalence of FS close to 1.3 and 1.5 suggests that these two common limits have an anchoring effect on predictions.

The problem required a lot of evidence to be integrated. Lommler (2012) suggested that engineers often struggle to integrate more than 2 to 3 key variables when making decisions. One possible reason for the poor predictions could then be that too much information was provided and participants were unable to distil the information within the time limit.

The exercise was carried out with a 30-minute count down timer on display and participants encourage to complete within the time limit. On average participants finished within 5 to 10 minutes of the time limit. It is well known that making decisions under time constraints significantly diminishes abilities (Kahneman 2012). One could sense participants getting nervous as the deadline approached and rushing to finish. Seventy six percent (76 %) of participants said they would need less than 60 more minutes to complete the task adequately. Five candidates said they would require a few days extra to discuss with colleagues, sleep on it or read through the literature to come to a decision. This latter group was probably more realistic in their assessment of their prediction.

A simple interpretation is that the bulk of participants were poor at making judgement calls. The middle panel in Figure 1 shows the experience of participants. The open symbols are the extreme values and the solid black lines illustrate the interquartile ranges. A gradual increase in experience occurs both to extremely low and extremely high likelihoods of failure. The higher levels of experience of participants that predicted failure is expected and reassuring. However, the high levels of experience of those predicting the chance of failure was extremely low is worrying.

The difficulty in using confidence in ones answer as a predictor of correctness is illustrated in the top panel of Figure 1. Confidence in answers increased towards the extreme ends of the probabilities (whether correct or incorrect). This highlights that verbal terms were also being used in a degree of belief sense. That is the more confidence a participant had in an outcome the lower or higher the stated probability. The fact that this occurred in both directions is worrying for expert-based assessments of risk. At the extremely low side two participants with more than 20 years of experience stated their confidence as 9 (corresponding to FS of 1.6 and 1.8). At the extremely high side two participants with more than 20 years of experience stated their confidence as 7 and 5 (both predicted FS = 0.7).

3.2 *Study 2*

Study 2 was devised to test the observation in Study 1 that the large amount of provided information appeared to result in poor judgements. Results from Study 2 showed that the amount of information relative to the number of controlling factors has an influence on judgement decisions.

To remove some ambiguity in interpreting results, participants were asked to select a likelihood of failure from a table that included both verbal and numerical expressions (Table 2). Nevertheless, deciding what is an acceptable probability of failure (PF) is not without its difficulties, considering the extensive discussion following Hendrik Kirsten's 1983 article "Significance of the probability of failure in slope engineering." Deciding on an acceptable probability of failure requires considering both the consequences of failure and the lifespan of a structure (CIRIA 1977). Consequently, a simple approach was taken to group responses into five categories. These five categories were: very low (PF ≤ 0.1, low (PF = 0.2), Even (PF = 0.5), high (PF = 0.8) and very high (PF ≥ 0.9) likelihood of failure. Combining several selections into extremely high and separately extremely low probabilities was because people struggle to understand extreme probabilities (Vick 2002).

Results to answers for the tailings questions are given in Figure 2a and Figure 2b. As this question was based on information for a tailings facility that had performed adequately, participants would be expected to give answers suggesting the facility had a small chance of failure. Results showed this to be the case, with on average for the paired questions, 36 % of responses being extremely low and 32 % being low. Answers suggesting the facility to be unsafe could be the result of conservatism and a realisation that whilst crucial information was provided large uncertainty remained. Considering the high experience of participants making these predictions, this may have been the case.

The influence of the amount of information on design questions were large uncertainty exists, is illustrated by comparing Figures 2a and 2b. Whilst one question (Fig. 2b) provided more detail than the other (Fig. 2a), the number of responses in the five categories were essentially the same. Thus, for this question with a large amount of epistemic uncertainty, providing a little more information had no real influence on results. The ratio of provided information did not materially change relative to the number of controlling factors.

It is evident from both Figures 2a and 2b that levels of experience (this time measured in terms of facilities) in the different categories was not substantially different. This suggests experience had little influence over decisions in this case.

The second question pair in Study 2 was on cover design. Whilst a number of factors control cover stability, slope angle and interface friction angle have the largest influence (a cover in its simplest form is an infinite slope). For the question where only slope angle was given (Fig. 2c) it would be expected that more respondents would state the design to be unsafe as liner interface friction angles can be lower than 10°. This was the case as the number of responses in the high to extremely high category in Figure 2c was double that in Figure 2d (where an interface friction angle was given).

Nevertheless, for the case where no interface friction angle was given, a large number of respondents predicted the cover would be stable (Fig. 2c). However, the respondents making these incorrect predictions tended to be less experienced than those making correct predictions.

The probability of failure for the case were an interface friction angle was given was calculated to be 0.14 (Koerner & Hwu 1991). Thus, the increase in extremely low and low responses when a friction angle was given (compare Fig. 2d to Fig. 2c) is a reflection of good judgement. However, this shift was predominantly within experienced respondents. Splitting responses between those with experience on 5 or less facilities and those with more, shows that this shift was amongst experienced respondents (Table 3).

Table 3. Possible selections for Study 2

Median probability of failure	Experienced	Inexperienced
Less detail (Unsafe)	0.8	0.2
More detail (Safe)	0.1	0.2

Considering the overall results for the cover questions it is evident that for this problem where less epistemic uncertainty existed, better judgements were made.

Further, giving more detail had a marked influence on the results. The ratio of provided information changed significantly relative to the number of controlling factors and this influenced predictions.

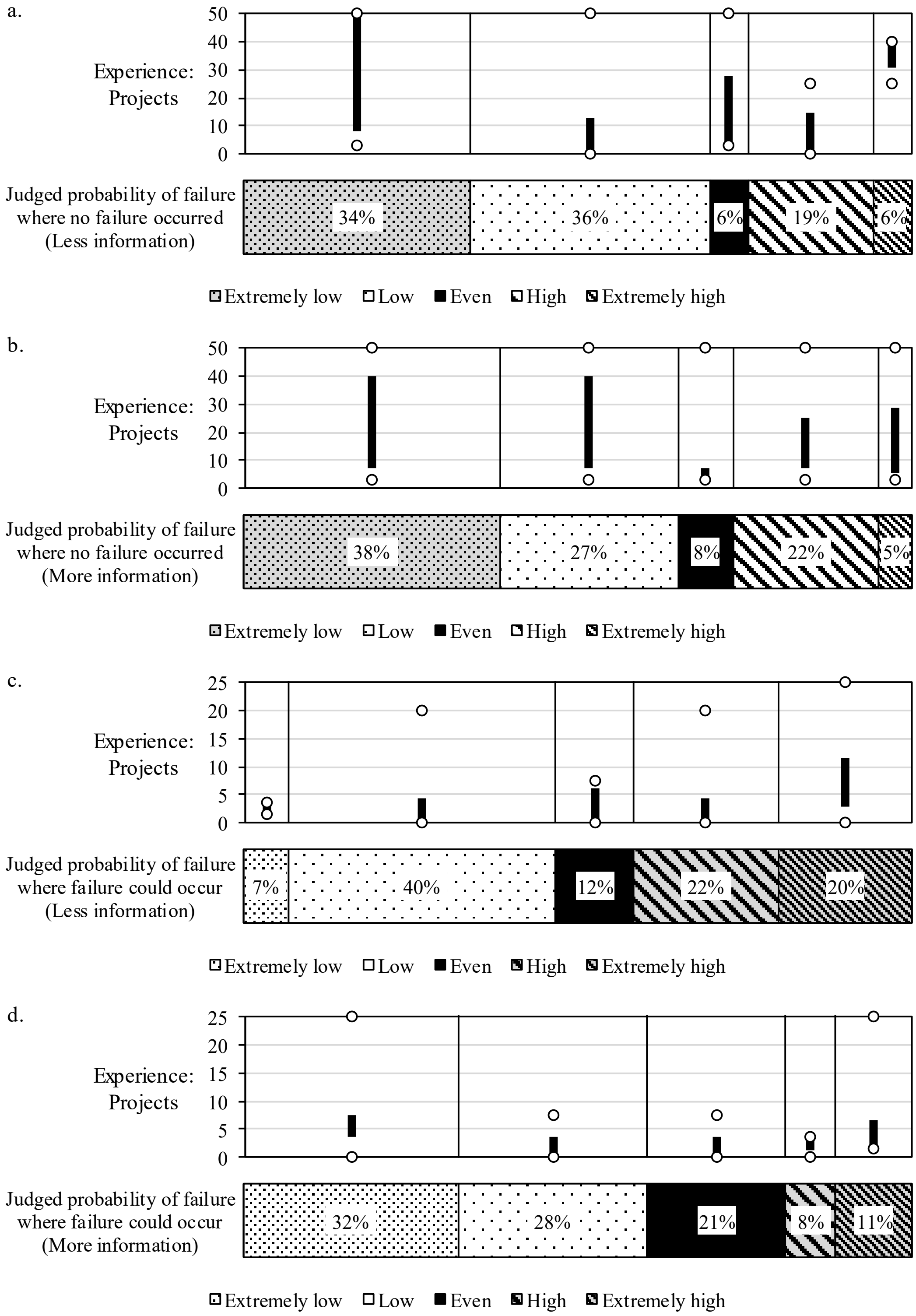

Figure 2. Results from Study 2: a. Tailings question with less information, b. Tailings question more information, c. Cover question less information, and d. Cover question more information (Sample size = 113, Group A = 53, Group B = 60)

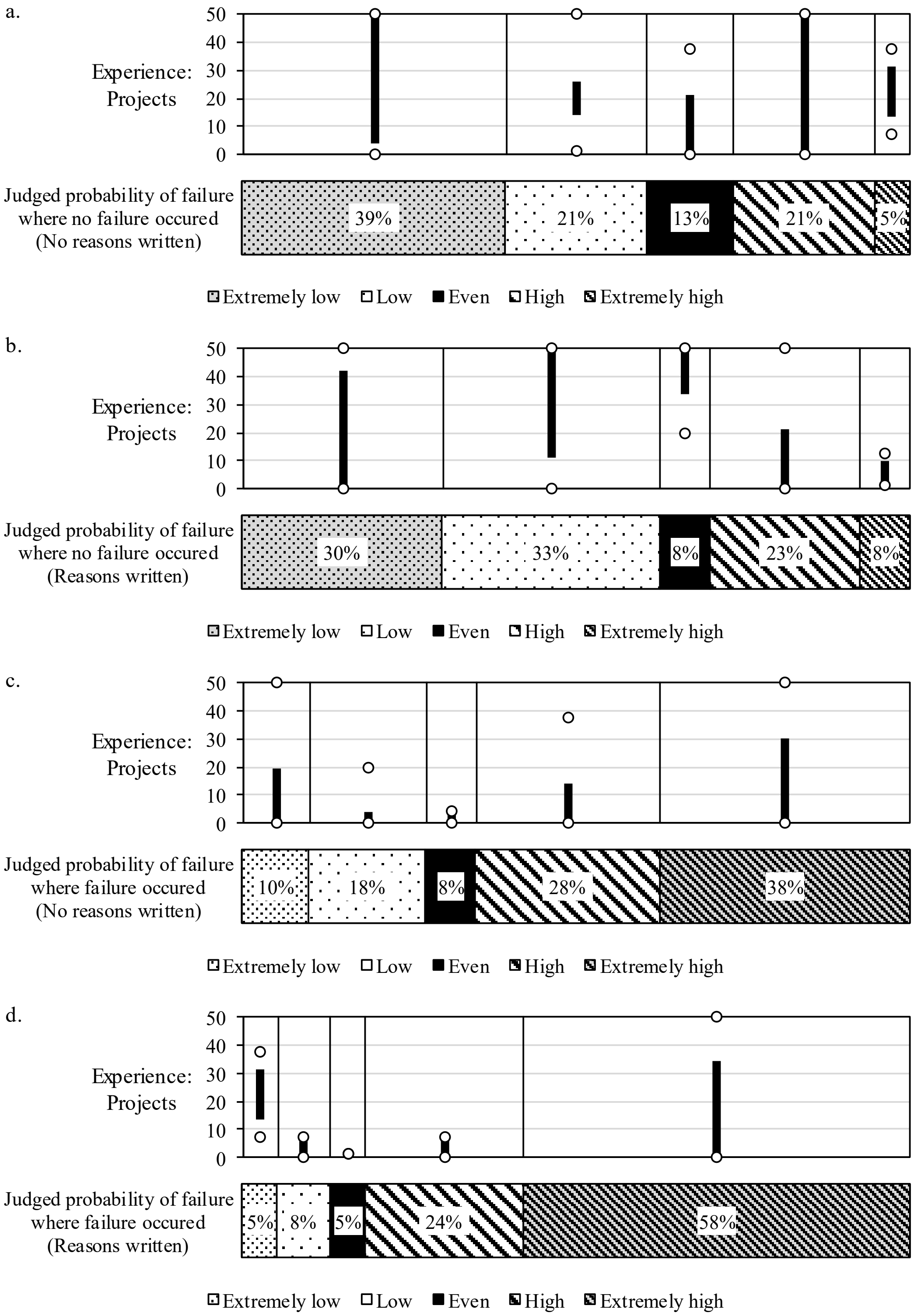

Figure 3. Results from Study 3: a. Foundation question no reasoning, b. Foundation question with reasoning, c. Gabion question no reasoning, and d. Gabion question with reasoning (Sample size = 78, Group A = 38, Group B = 40)

3.3 *Study 3*

As Study 2 showed that engineering judgement is better applied to problems with less controlling variables, similar problems were selected for Study 3. Two questions were presented: one on a foundation, and another on a gabion wall. The study then looked to see whether removing time constraints and allowing participants to detail reasoning would improve predictions. This study used the same likelihood selections as Study 2 and so the same categorisation was used to analyse results.

Although the loads on the foundation were high, the applied pressure was reasonably low at 115 kPa, the eccentricity was well within the pads middle third (width = 16.5 m and eccentricity = 2.1 m) and the soil consistency was medium dense to dense. The likelihood of complete foundation failure should therefore be seen as low to extremely low, particularly as complete failures of foundations are less prevalent than settlement failures.

The majority of responses were low to extremely low and documenting decision reasons had a minimal influence (Figure 3a and Figure 3b). The minimal influence of documenting reasoning may be a consequence of the relative simplicity of the problem. Experience (measured in terms of projects) did not have a significant influence on results. The scatter in results between categories may reflect a poor understanding of probabilities of failure for foundations were factors of safety have traditionally been used. Alternatively, the scatter could reflect conservatism.

Results for the gabion wall question are shown in Figure 3c and 3d. Whilst a number of factors influence the performance of gabion walls (Agostini et al. 1987) this 4 m high gabion wall was built on a 2.8 m thick clay layer with an SPT N value of 3 (all information provided to participants). Such a low SPT N value in clay should raise serious concerns for a geotechnical practitioner.

The majority of responses were in the high to extremely high categories, however documenting reasons resulted in a significant increase in responses within this category. Due to documenting reasoning the number of responses categorised as extremely high increasing from 38% (Figure 3c) to 58% (Figure 3d). Identifying the clay layer was crucial to making a correct prediction. Spending extra time documenting reasoning resulted in a larger proportion of participants making a correct assessment. The fact that so many predicted failure was unlikely, when not asked to state reasoning, highlights the importance of such a simple step to judgement.

Comparing poor predictions of failure in Study 1 to good predictions in Study 3, highlights the importance of limiting judgement calls to problems with few controlling factors and removing time constraints. Similar to Study 1, responses at the extreme ends of likelihood were also made by respondents with increasing experience in Study 3. As before the higher levels of experience of those predicting failure is expected and reassuring. Concern is also raised at the high levels of experience of those predicting extremely low probabilities of failure especially for this simpler problem.

4 CONCLUSIONS

Does judgement work and how can we improve it? The studies presented suggest that for simple problems with a few controlling factors the majority of practitioners could make good calls. Predicting who these individuals are is a difficult task. Experience in terms of years or projects/facilities worked on seems a poor predictor. Where any trend with experience was evident, it tended to suggest experience leads to both extremely good and extremely poor predictions. Self-confidence was an even weaker predictor. One valuable conclusion then appears: judgement is fallible despite experience or confidence.

The following are some recommendations for improving one's judgement:

- Beware of anchor values. Acceptable factors of safety or probabilities of failure can cause us to overlook key information in order to obtain answers close to what is expected.
- Beware of variables. Problems with more than 2 to 3 controlling factors are difficult and need to be broken down into simpler steps. Judgement cannot be used as a shortcut to a solution.
- Beware of time constraints. Working under time pressures encourages shortcuts and these will result in judgement errors.
- Beware of judging first. Spending a few minutes writing or sketching can prevent key factors being overlooked. Finally, when you reach a conclusion make sure you haven't overlooked other potential solutions.

5 ACKNOWLEDGMENTS

The author wishes to acknowledge all the participants who made time in their busy schedules to take part in these studies. Comments by Dr Nico de Koker (Stellenbosch University) greatly improved the quality of this paper. The assistance of Dr Florian Hortkorn (Jones and Wagener) and Dr Irvin Luker (University of Witwatersrand) in developing the questions is appreciated.

6 REFERENCES

Agostini, R. Cesario, L. Conte, A. Masetti, M. & Papetti, A. 1987. Flexible gabion structure in earth retaining works, Officine Maccaferri S.p.A., Bologna, Italy.

Azam, S. & Li, Q. 2010. Tailings dam failures: a review of the last one hundred years. *Geotechnical News*. 28(4): 50-54.

Blight, G.E. 2010. *Geotechnical engineering for mine waste storage facilities*. Taylor & Francis Group: London, UK.

CIRIA. 1977. Rationalisation of safety and serviceability factors in structural codes, CIRIA, London.

Davie, J.R. & Lewis, M.R. 1988. Settlement of Two Tall Chimney Foundations. In *Second International Conference on Case Histories in Geotechnical Engineering, S. Prakash, ed., University of Missouri-Rolla, St. Louis*: 1309-1313.

De Mello, V.F. 1977. Reflections on design decisions of practical significance to embankment dams. *Géotechnique*. 27(3): 281-355.

Kahneman, D. 2012. *Thinking, fast and slow*. Penguin Books: United Kingdom.

Kahneman, D. & Klein, G. 2009. Conditions for intuitive expertise: a failure to disagree. *American psychologist*. 64(6): 515.

Koen, B.V. 2003. *Discussion of the method: Conducting the engineer's approach to problem solving*. Oxford University Press on Demand: New York.

Koerner, R.M. & Hwu, B.-L. 1991. Stability and tension considerations regarding cover soils on geomembrane lined slopes. *Geotextiles and Geomembranes*. 10(4): 335-355.

Lommler, J.C. 2012. *Geotechnical problem solving*. John Wiley & Sons.

MacRobert, C.J. 2018a. Factors of safety and probabilities of failure in geotechnical engineering: What do we mean? *Civil Engineering, SAICE*: 45-50.

MacRobert, C.J. 2018b. Slope stability: overconfidence in experts and novices. In *6th International Mining and Industrial Waste Management Conference, C. J. MacRobert, ed., SAICE Geotechnical Division, Legend Golf and Safari Resort, Limpopo, South Africa*: 99-108.

MacRobert, C.J. & Blight, G.E. 2013. A field study of the in situ moisture regime during active hydraulic tailings deposition. *Journal of the South African Institution of Civil Engineering*. 55(3): 57-68.

Nortje, R.A. 2018. Capping of mining and industrial waste management facilities in South Africa: What is required? In *6th International Mining and Industrial Waste Management Conference, C. J. MacRobert, ed., SAICE Geotechnical Division, Limpopo, South Africa.*

Nowatzki, E.A. & Wrench, B.P. 1988. Geotechnical Investigation into Causes of Failure of a Gabion Retaining Wall. In *Second International Conference on Case Histories in Geotechnical Engineering, S. Prakash, ed., University of Missouri-Rolla, St. Louis*: 1477-1481.

Petroski, H. 1985. *To engineer is human*. Vintage Books: New York.

Reagan, R.T. Mosteller, F. & Youtz, C. 1989. Quantitative meanings of verbal probability expressions. *Journal of Applied Psychology*. 74(3): 433-442.

Robinson, B. 2008. Fraser Alexander tailings disposal course, Fraser Alexander, Witfield, South Africa.

Samuelson, P.A. 1974. Challenge to judgment. *The Journal of Portfolio Management*. 1(1): 17-19.

Sherif, M. Taub, D. & Hovland, C.I. 1958. Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of experimental psychology*. 55(2): 150.

Sowers, G.F. 1993. Human factors in civil and geotechnical engineering failures. *Journal of Geotechnical Engineering*. 119(2): 238-256.

USBR-USACE. 2015. Best Practices in Dam and Levee Safety Risk Analysis, U.S. Bureau of Reclamation - U.S. Army Corps of Engineers.

Vick, S.G. 2002. *Degrees of belief: Subjective probability and engineering judgment*. ASCE Publications: Reston, Virginia.